Theor Appl Genet (2010) 121:1209–1225
DOI 10.1007/s00122-010-1382-x

ORIGINAL PAPER

Genetic diversity and linkage disequilibrium studies on a 3.1-Mb genomic region of chromosome 3B in European and Asian bread wheat (*Triticum aestivum* L.) populations

C. Y. Hao · M. R. Perretant · F. Choulet · L. F. Wang · E. Paux · P. Sourdille · X. Y. Zhang · C. Feuillet · Francois Balfourier

Received: 11 December 2009 / Accepted: 3 June 2010 / Published online: 18 June 2010

Abstract Genetic diversity and linkage disequilibrium (LD) were investigated in 376 Asian and European accessions of bread wheat (*Triticum aestivum* L.). After a first and rapid screening about diversity and genetic structure at the whole genome scale using 70 simple sequence repeats (SSRs), we focused on a sequenced contig (*ctg954*) of 3.1 Mb located on the short arm of chromosome 3B of cv. Chinese Spring, using 32 SSRs and 10 single nucleotide polymorphisms. This contig is part of a multiple fungal resistance region. Mean polymorphism information content value on the 32 SSRs was slightly higher in the Asian genepool (0.396) than that for the European (0.329) pool. Compared with results at the whole genome scale, data from this 3.1-Mb region indicated similar trends in genetic diversity indices between both genepools. Population structure and molecular variance analyses demonstrated significant genetic differentiation and geographical subdivision in both groups of accessions. Concerning LD at the contig level, the European population had a significantly higher mean r^2 value (0.23) than the Asian population (0.18), indicating a stronger LD in the European material. With a mean of 1 marker every 74 kb, the resolution reached here allowed to perform a detailed comparative analysis of the LD and genetic diversity along the complete 3.1-Mb region in both genepools. A sliding-window approach revealed some interesting regions of the contig where LD is increasing when genetic diversity is decreasing. This study provides an in-depth understanding of molecular population genetics in European and Asian wheat gene pools, and prospects for association mapping of important sources of fungal disease resistance.

Communicated by A. Graner.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1382-x) contains supplementary material, which is available to authorized users.

C. Y. Hao · M. R. Perretant · F. Choulet · L. F. Wang · E. Paux · P. Sourdille · C. Feuillet · F. Balfourier (✉)
INRA UMR 1095,
Genetics Diversity and Ecophysiology of Cereals,
234 Avenue du Brezet, 63100 Clermont-Ferrand, France
e-mail: balfour@clermont.inra.fr

C. Y. Hao · L. F. Wang · X. Y. Zhang
Key Laboratory of Crop Germplasm Resources and Utilization,
Ministry of Agriculture, CAAS, Beijing 100081, China

C. Y. Hao · L. F. Wang · X. Y. Zhang
The National Key Facility for Crop
Gene Resources and Genetic Improvement,
CAAS, Beijing 100081, China

C. Y. Hao · L. F. Wang · X. Y. Zhang
Institute of Crop Science, CAAS, Beijing 100081, China

Introduction

Bread or common wheat (*Triticum aestivum* L.) is one of the most important cereal crops worldwide, taking third place in world production after maize and rice, and covering the largest area among cereals (217 million ha), mainly in Europe and Asia [57 and 100 million ha, respectively (http://faostat.fao.org/default.aspx)]. Domestication of bread wheat was a complex process. It is accepted that *T. aestivum* originated from a cross (crosses) between domesticated hulled tetraploid emmer *T. dicoccon* and the goat grass *Aegilops tauschii* (DD). This cross should have taken place after emmer (or hard wheat) cultivation spread east from the Fertile Crescent into the natural distribution area of *A. tauschii*. The cross occurred most probably South or West of the Caspian Sea about 8,000 years ago (Kilian et al. 2009). So,

domestication of bread wheat can be dated back to 8,000 BC in this part of the world as indicated by archeological evidence (reviewed by Feldman 2000). Then, cultivated bread wheats were presumably spread both westwards (to Europe) and eastwards (to Asia) from the Fertile Crescent by early farmers (Schlumbaum et al. 1998; Nesbitt 2001; Balfourier et al. 2007; Mitrofanova et al. 2009). Natural adaptation to different climatic and environmental conditions, and selection practices undertaken by farmers to locally preserve favored genotypes, generated much of the variability currently used by plant breeders.

DNA-based markers, recently compared by Able et al. (2007), such as simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs), can be used to characterize genetic resources, map plant genomes, and tag genomic regions and/or specific genes (Dreher et al. 2003). Once DNA markers are shown to be associated with a target trait, plant breeders can use them to identify individual plants in large segregating populations with greater probabilities of phenotypes containing particular alleles of interest (Tanksley et al. 1989; Morris et al. 2003). Genetic diversity is a key factor enabling adaptation of natural populations in changing environments (Väli et al. 2008); evaluations of genetic diversity are common in population genetics, and are particularly important in conservation genetics (Frankham 2005). In recent years, many studies focused on the levels of genetic diversity in European (Huang et al. 2002; Röder et al. 2002; Roussel et al. 2005; Balfourier et al. 2007), and Asian wheat genepools (Ghimire et al. 2005; Hao et al. 2006, 2008). Independently, these studies provided a considerable overall understanding of the genetic diversity levels in both groups. Most part of diversity analyses on wheat were assessed at the whole genome scale, with the exception of recent study at the 3B chromosome scale (Horvath et al. 2009). Linkage disequilibrium (LD), or nonrandom associations of alleles at adjacent loci within a population, is becoming increasingly important for its potential effect on association mapping strategies in natural populations (Wilson et al. 2004). Numerous forces including domestication processes, population subdivision and selective breeding can increase LD throughout the genome or in genomic segments flanking selected loci (Rafalski and Morgante 2004). Bread wheat, a self-pollinating species, has undergone continual domestication and strong selection pressures throughout its history; this can create population bottlenecks, decrease genetic diversity and increase LD. To date, the extent of LD patterns in bread wheat and durum wheat (*T. turgidum* L. ssp. *durum*) have been investigated on both genome-wide or on specific chromosomal regions (Maccaferri et al. 2005; Breseghello and Sorrells 2006; Chao et al. 2007; Somers et al. 2007; Raquin et al. 2008; Horvath et al. 2009). For example, Breseghello and Sorrells (2006) described that consistent extent of LD was less than 1 cM on chromosome 2D while about 5 cM in the centromeric region of 5A using 33 and 20 SSR markers, respectively. Analyzing diversity of 242 whole genome SSRs among 43 US wheat elite cultivars, Chao et al. (2007) revealed that genome-wide LD estimates were generally less than 1 cM for the genetically linked loci pairs and that most of the LD regions observed were between loci less than 10 cM apart. Somers et al. (2007) suggested that LD mapping of wheat can be performed with SSRs to a resolution of <5 cM. However, despite the relatively large number of studies for this crop, there is still inadequate information about the extent and the pattern of LD at the scale of a BAC contig. This may be explained by the delay in building physical map and a lack of high density markers due to a huge genome size, as well as difficulty in finding locus-specific markers due to polyploidy.

In the present study, our objectives were to evaluate and compare the levels and patterns of both LD and molecular diversity among 376 wheat accessions representing the diversity of European and Asian wheat genepools. After a first and rapid screening about diversity and genetic structure at the whole genome scale, we focused on a BAC contig of 3.1 Mb from the physical map of chromosome 3B (*ctg954*, Paux et al. 2008). This contig of 26 BACs is part of a multiple fungal disease resistance region (*R* locus), encompassing the putative *Fhb1* locus of resistance to Fusarium head blight (FHB). From this contig, we genotyped 42 linked loci in the two genepools to study the relationships between gene density, genetic diversity and LD at the scale of 3 Mb-sized sequenced region. The ultimate aim was to gain important information to allow further association analysis and novel resistance gene discovery for wheat improvement applications.

Materials and methods

Plant material

Representative collections of bread wheat from a wide range of geographical areas in Europe and Asia (East Asia and South Asia) were obtained from the Clermont-Ferrand Genetic Resources Center (http://www.clermont.inra.fr/umr-asp). The two genepools, European and Asian, comprised 21 and 20 populations, respectively. These 41 populations are stemming from 34 different countries. From these countries, China is represented by 15 populations, including 22 provinces with different ecological environments (see figure in supplementary material). The populations were constructed using preliminary studies of genetic diversity in the same geographical areas (Wang et al. 2007). In total, 189 accessions from European

countries, and 187 accessions from East Asia were used in the present study, detailed information on these 376 accessions are given in supplementary Table 1.

Seedlings of the 376 accessions were germinated and grown in the greenhouse. After 3–4 weeks, 30–50 mg of fresh leaf tissue per accession were placed in 1.2-ml collection microtubes with 3–5 stainless steel beads (3 mm diameter) in a 96-well plate and frozen in liquid nitrogen. The frozen leaf samples were thoroughly disrupted with an industrial shaker, and genomic DNA was purified using a BioSprint 96 DNA Plant Kit (QIAGEN Ltd, West Sussex, UK) according to the BioSprint 96 User Manual (http://www.qiagen.com).

Molecular marker analysis

Two different sets of molecular markers were used in the present study:

- To assess the genetic diversity and to estimate the genetic structure at the whole genome scale in the whole sample of 376 accessions, a first set of 70 molecular markers [GWM genomic-SSR (Röder et al. 1998), GPW (Nicot et al. 2004) and CFE EST-SSRs markers (Zhang et al. 2005)] were selected in order to give good coverage of the whole genome map. Chromosome location of these 70 SSR locus is given in Table 1, while sequences, melting temperature and details of primers are given in the GrainGenes database (http://wheat.pw.usda.gov).
- To study the genetic diversity and to estimate LD at the scale of a Mb-sized sequenced region, a second set of 42 molecular markers were developed from the *ctg954* contig (Genebank accession number: FN564434). This contig of 26 BACs (see supplementary Table 2) was entirely sequenced. Annotation revealed 53 genic regions (Choulet et al. 2010) comprised of 41 genes likely to be functional (with a complete ORF) and 12 pseudogenes and truncated genes fragments that are not under selection (see supplementary Table 3). From the complete sequence, we defined and selected 32 CFB genomic-SSRs markers showing a good amplification profile, a high level of polymorphism and a relative even distribution along the sequence. Information regarding primer sequences, motifs, expected sizes and physical positions of all CFB markers is provided in supplementary Table 4. Then, to fill some gaps remaining in the SSR distribution we added 10 insertion site-based polymorphism (ISBP)-derived SNPs markers (Paux et al. 2010) (namely CFP markers in supplementary Table 5). The total of 42 markers is scattered along the contig with a density of 1 marker for 74 kb (73.8 ± 16 kb).

Genotyping for all SSRs by PCR was conducted using M13-marked primers as described in Saintenac et al. (2009). Each locus was amplified separately and then individual PCR products (up to 4) were multiplexed into panels by pooling on the basis of product size range and fluorescent label. Amplification products were visualized using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems; http://www.appliedbiosystems.com). Fragment sizes were calculated using GeneMapper v3.7 software (Applied Biosystems), where different alleles are represented by different amplification sizes of tandem repeats.

SNP genotyping was performed with the SNaPshot assay (Applied Biosytems), which involves single base extension (SBE) of an unlabeled primer designed to anneal one base upstream of the relevant SNP, followed by extension with a fluorescent-labeled dideoxy terminator. Primers for the SBE reaction were designed for different SNP sites after comparing sequencing data of eight bread wheat varieties (for primer sequences, see Table 3 in supplementary material). Five microliters of PCR products obtained through amplification of ISBP primers (for PCR reaction conditions, see Paux et al. 2006), to be used as template DNA of the SNaPshot PCR, were incubated at 37°C for 60 min with 1 U of shrimp alkaline phosphatase (SAP, GE Healthcare) and 0.1 U of exonuclease I (*Exo*I, Ozyme) to remove primer excess and dNTPs, followed by enzyme inactivation by heating at 80°C for 10 min. The SNaPshot SBE reaction was carried out in 10 µl containing 0.75 µl SNaPshot Reaction Mix, 3 µl of SAP/*Exo*I-treated PCR products, 1 µl PCR primer (2 µM), and 5.25 µl of Milli-Q water. Extension was performed for 25 cycles of denaturation at 96°C for 10 s, annealing at 50°C for 5 s, and extension at 60°C for 30 s. The extension products were treated with 1 U of SAP (GE Healthcare; http://www.gehealthcar.com) at 37°C for 60 min and 80°C for 10 min to remove unincorporated ddNTPs. For both clean-up reaction the specific buffer is the 10× buffer provided for Sap used at its accurate concentration. Two microliters of diluted extension products (10% in Milli-Q water) were combined with 9 µl of Hi-Di formamide (Applied Biosytems) and 0.09 µl of internal size standard GeneScan-120 LIZ^{TM} (Applied Biosytems). After denaturation, samples were run on an ABI PRISM 3100 Genetic Analyzer using the POP-7 polymer (Applied Biosytems). Data were analyzed using GeneMapper v3.7 software (Applied Biosystems) with the peak threshold set to a minimum of 100 relative fluorescence units because genotyping using the SNaPshot assay depends on fluorescence signal intensity to identify the SNP alleles.

Table 1 Diversity indices of European and Asian accessions using 70 genome-wide SSRs

SSR locus	Chrom. location	Total no. of alleles			No. private alleles		PIC value		
		Overall	Europe	Asia	Europe	Asia	Overall	Europe	Asia
Xgpw7072	1AS	7	7	6	1	0	0.758	0.691	0.709
Xgwm99	1AL	10	8	9	1	2	0.647	0.447	0.715
Xgwm135	1AL	25	20	14	11	5	0.756	0.715	0.780
Xgwm11	1 BS	13	12	12	1		0.805	0.792	0.809
Xgwm413	1 BS	16	15	11	5		0.860	0.756	0.865
Xgpw7443	1BL	3	2	3	0		0.367	0.064	0.528
Xgpw7577	1BL	5	4	5	0		0.250	0.136	0.352
Xgwm337	1DS	18	17	13	5		0.864	0.795	0.863
Xgpw7082	1DS	4	3	3	1		0.170	0.254	0.066
Xgwm642	1 DL	7	6	5	2		0.624	0.644	0.589
Xcfe175	2A	3	3	3	0	0	0.432	0.330	0.501
Xgwm372	2AS	28	20	28	0	8	0.917	0.891	0.913
Xgwm312	2AL	33	19	26	7	14	0.883	0.842	0.879
Xgpw7570	2AL	4	4	2	2	0	0.058	0.104	0.011
Xcfe52	2B	12	7	9	3	5	0.517	0.430	0.587
Xgwm257	2BS	6	6	3	3	0	0.643	0.560	0.610
Xgpw7438	2BS	2	2	2	0	0	0.500	0.500	0.500
Xgwm120	2BL	19	17	16	3	2	0.872	0.854	0.785
Xcfe68	2D	2	2	2	0	0	0.136	0.236	0.022
Xgwm261	2DS	18	12	17	1	6	0.800	0.711	0.808
Xgwm539	2DL	29	20	28	1	9	0.898	0.875	0.901
Xgwm2	3AS	8	6	8	0	2	0.630	0.404	0.740
Xgwm480	3AL	11	5	10	1	6	0.541	0.212	0.753
Xgpw7213	3AL	3	3	3	0	0	0.295	0.064	0.462
Xcfe172	3D	3	3	2	1	0	0.138	0.160	0.112
Xgwm566	3BS	11	7	11	0	4	0.807	0.738	0.817
Xgpw7452	3BS	8	7	5	3	1	0.585	0.638	0.453
Xgwm341	3DS	20	16	17	3	4	0.880	0.854	0.864
Xgwm664	3DL	6	3	6	0	3	0.368	0.197	0.472
Xcfe300	4A	9	6	8	1	3	0.637	0.661	0.504
Xgwm610	4AS	15	8	12	3	7	0.686	0.555	0.747
Xcfd71a	4AL	9	5	8	1	4	0.421	0.324	0.498
Xcfe8	4BS	4	3	4	0	1	0.669	0.636	0.658
Xgwm251	4BL	17	14	14	3	3	0.825	0.821	0.788
Xgwm149	4BL	10	6	10	0	4	0.609	0.520	0.684
Xgpw7241	4BL	2	2	2	0	0	0.033	0.033	0.033
Xgpw7795	4DS	7	2	6	1	5	0.074	0.072	0.074
Xcfd71d	4DL	18	15	17	1	3	0.896	0.846	0.893
Xcfe186	5A	3	3	3	0	0	0.450	0.528	0.310
Xgwm415	5AS	6	5	5	1	1	0.592	0.624	0.523
Xgpw7218	5AS	4	4	4	0	0	0.607	0.536	0.616
Xgwm186	5AL	25	16	23	2	9	0.898	0.804	0.856
Xgwm234	5BS	11	10	11	0	1	0.850	0.801	0.823
Xgwm408	5BL	16	12	14	2	4	0.849	0.761	0.850
Xgpw7425	5BL	2	2	2	0	0	0.196	0.141	0.246
Xcfe301	5D	6	6	5	1	0	0.610	0.559	0.649
Xgwm190	5DS	16	9	15	1	7	0.711	0.652	0.746

Table 1 continued

SSR locus	Chrom. location	Total no. of alleles			No. private alleles		PIC value		
		Overall	Europe	Asia	Europe	Asia	Overall	Europe	Asia
Xgwm272	5DL	7	7	6	1	0	0.654	0.656	0.642
Xgpw7592	6AS	2	2	2	0	0	0.283	0.168	0.381
Xcfe273	6AS	2	2	2	0	0	0.311	0.184	0.409
Xgwm427	6AL	18	15	16	2	3	0.834	0.845	0.803
Xgpw7384	6AL	4	4	3	1	0	0.312	0.213	0.394
Xgwm626	6BS	9	8	7	2	1	0.502	0.521	0.474
Xgwm219	6BL	20	17	15	5	3	0.850	0.850	0.821
Xcfe95	6D	2	2	1	1	0	0.011	0.022	0.000
Xgwm469	6DS	14	12	12	2	2	0.838	0.838	0.816
Xgwm325	6DS	8	8	8	0	0	0.795	0.686	0.839
Xgpw7433	6DL	3	3	2	1	0	0.386	0.135	0.494
Xgwm260	7AS	21	18	16	5	3	0.839	0.768	0.875
Xgpw7185	7AS	3	3	2	1	0	0.058	0.032	0.084
Xgpw7288	7AS	9	7	6	3	2	0.704	0.613	0.745
Xgpw7386	7AL	9	8	8	1	1	0.788	0.746	0.786
Xgwm400	7BS	17	14	13	4	3	0.859	0.777	0.816
Xgwm46	7BS	21	17	16	5	4	0.890	0.847	0.871
Xgpw7320	7BS	12	8	11	1	4	0.792	0.644	0.789
Xgpw7342	7BS	10	8	10	0	2	0.755	0.687	0.746
Xgpw7596	7BL	2	2	2	0	0	0.274	0.186	0.348
Xgwm44	7D	14	12	13	1	2	0.879	0.852	0.880
Xcfe135	7D	3	3	2	1	0	0.290	0.365	0.198
Xgwm437	7DL	19	14	19	0	5	0.870	0.806	0.895
Total		743	578	634	109	165			
Mean ± SE		10.6 ± 0.9	8.3 ± 0.7	9.1 ± 0.8			0.591 ± 0.032	0.532 ± 0.034	0.601 ± 0.032

Statistical analyses

Diversity analysis

Standard statistics of genetic diversity, including total number of alleles and polymorphism information content (PIC value) at each SSR locus, were carried out with PowerMarker v3.25 (Liu and Muse 2005). PIC value was calculated according to the formula $H = 1 - \sum p_i^2$ (Nei 1973), where p_i means the frequency of the ith allele. These indices were determined on the two different sets of SSR markers (70 SSRS from the whole genome and 32 SSRs from ctg954 BACs contig). To calculate these indices, we used the two basic genepools (Europe and Asia) as well as the 41 populations of different geographical origin in order to make comparisons at different scales.

The genetic variation within and among populations of wheat accessions was tested using analysis of molecular variance (AMOVA) implemented in Arlequin v3.11 (Excoffier et al. 2005).

Population structure analysis

The knowledge of genetic structure among a panel of accessions is an important stage to understand and compare whole genetic diversity. For further association analyses, in order to avoid spurious association, it may be also useful to know structure components to be incorporated as covariates in association tests. To be powerful, population structure analysis must be investigated at the widest genome scale. So, to infer the population structure in the present set of accessions with STRUCTURE v2.2 software (Pritchard et al. 2000), we only considered data from the 70 genome-wide microsatellite loci. We adopted the "admixture model", length of burn-in period equal to 50,000 iterations and a run of 100,000 replications of Markov Chain Monte Carlo (MCMC) after burn in. For each run, 5 independent runs of STRUCTURE were performed with the number of cluster K varying from 2 to 10, leading to 45 Structure outputs. Then we estimated the number of founding populations and the best output on the

basis of the Evanno criterion (2005). In order to confirm the result of the population structure obtained from STRUCTURE, individuals were assigned to the 41 populations according to their geographical origins. Then, individual allelic frequencies of the 70 SSRs covering the whole genome were calculated for each 41 population. The frequency matrix formed by all populations was then used to compute Manhattan dissimilarity coefficients between pairs of populations, and finally, a neighbor-joining tree was created from the Manhattan dissimilarity matrix using DARwin v5 software (Perrier et al. 2003).

Linkage disequilibrium analysis along ctg954

The pairwise estimates squared allele-frequency correlations (r^2), and significance of each pair of loci (Gaut and Long 2003) were performed using the 32 polymorphic SSRs and 10 SNPs spanning an approximate 3.1-Mb region on chromosome 3B, with the dedicated procedure of the TASSEL software (Zhang et al. 2006), using 100,000 permutations. To describe the relationship between the linkage disequilibrium decay and genetic diversity along the chromosomal region, and to avoid variations in the estimates of LD among any pairs of markers caused by decreasing disequilibrium with increasing physical distance, we used a sliding window approach with $n = 5$ markers per window.

Results

Marker design and genetic diversity in European and Asian wheat germplasm

When considering the first set of 70 SSR at the whole genome scale, a total of 578 and 634 alleles were detected in the European and Asian genepools, respectively (Table 1). Due to high differences between loci for their polymorphism level, there was not significant difference between the two genepools in terms of mean number of alleles per locus. However, if we define a private allele as an allele which is only present in one of the two genepools, 109 and 165 alleles may be considered as private alleles in the European and Asian genepools, respectively. Compared to the overall number of alleles (743 alleles), more than one-third of the observed alleles were not shared by the two genepools. The average Asian PIC value (0.601) was relatively higher ($P > 0.05$) than that for the European genepool (0.532). So globally, at the genome scale, the two genepools significantly differ for both diversity indices.

At the level of 42 loci along the ctg954 contig, a total of 127 and 131 alleles were detected in the European and Asian genepools, respectively (Table 2). For these 42 loci, there was no significant difference between the two genepools in terms of total numbers of alleles. Among the 42 loci, 34 loci had the same alleles between two groups, but the frequency differed for these alleles. Once again, the average Asian PIC value (0.396) appeared relatively higher ($P > 0.05$) than that for the European genepool (0.329). For 14 of the 42 tested loci, the PIC value was significantly different between Asian and European genepools. A total of 8 loci among the 42 tested loci showed private alleles: Xcfb6015 (2 alleles), Xcfb6017 (1 allele), Xcfb6046 (1 allele) and Xcfb6047 (1 allele) for Europe; Xcfb6072 (1 allele), Xcfb6047 (1 allele), Xcfb6058 (5 alleles), Xcfb6061 (1 allele) and the same Xcfb6046 (1 allele) for Asia. These eight loci are evenly distributed along the 3.1 Mb.

Furthermore, the structure of genetic diversity based on the 41 different geographical origins calculated by AMOVA (Table 3) revealed that within-population differences accounted for 73% of the total genetic variation ($P < 0.001$) whereas 21% were attributed to differences among the 41 populations ($P < 0.001$) and, finally, only

Table 2 Diversity indices of European and Asian accessions using 42 markers from *ctg954*

Markers	Number of allele			PIC value		
	Overall	Europe	Asia	Overall	Europe	Asia
cfb6092	3	3	3	0.551	0.507	0.566
cfp5005_S170[a]	2	2	2	0.421	0.319	0.486
cfb6026	2	2	2	0.345	0.420	0.235
cfp5009_S142[a]	2	2	2	0.455	0.426	0.478
cfb6015	7	7	5	0.762	0.763	0.737
cfb6096	2	2	2	0.223	0.134	0.310
cfb6031	9	9	9	0.808	0.773	0.800
cfb6032	2	2	2	0.443	0.484	0.346
cfp5019_S73[a]	2	2	2	0.344	0.271	0.404
cfb6022	2	2	2	0.341	0.268	0.402

Table 2 continued

Markers	Number of allele			PIC value		
	Overall	Europe	Asia	Overall	Europe	Asia
cfb6033	3	3	3	0.511	0.422	0.568
cfb6101	2	2	2	0.457	0.469	0.441
cfp5022_S190[a]	2	2	2	0.435	0.376	0.475
cfb6035	2	2	2	0.065	0.056	0.075
cfb6105	2	2	2	0.047	0.011	0.082
cfp5026_S67[a]	2	2	2	0.493	0.428	0.489
cfb6017	6	6	5	0.469	0.544	0.353
cfb6107	4	4	4	0.558	0.477	0.574
cfp5028_S91[a]	2	2	2	0.496	0.458	0.492
cfp5029_S127[a]	2	2	2	0.436	0.347	0.489
cfb6109	2	2	2	0.079	0.123	0.033
cfb6040	2	2	2	0.167	0.246	0.074
cfb6110	2	2	2	0.438	0.447	0.429
cfb6043	2	2	2	0.053	0.069	0.039
cfb6072	6	5	6	0.614	0.589	0.628
cfb6045	5	5	5	0.544	0.513	0.564
cfb6046	7	6	6	0.543	0.418	0.631
cfp5046_S123[a]	2	2	2	0.444	0.454	0.433
cfb6047	4	3	4	0.167	0.201	0.128
cfb6078	2	2	2	0.376	0.245	0.461
cfb6053	3	3	3	0.140	0.078	0.199
cfb6055	2	2	2	0.439	0.311	0.497
cfb6056	2	2	2	0.321	0.129	0.446
cfb6057	4	4	3	0.257	0.243	0.266
cfb6058	12	7	12	0.485	0.321	0.626
cfb6059	4	4	4	0.320	0.297	0.338
cfb6067	2	2	2	0.089	0.093	0.085
cfp5061_S154[a]	2	2	2	0.258	0.282	0.233
cfp5062_S38[a]	2	2	2	0.394	0.275	0.471
cfb6061	4	3	4	0.488	0.310	0.610
cfb6011	2	2	2	0.148	0.124	0.172
cfb6012	3	3	3	0.304	0.100	0.458
Total	136	127	131			
Mean ± SE	3.2 ± 0.3	3.0 ± 0.3	3.1 ± 0.3	0.375 ± 0.028	0.329 ± 0.028	0.396 ± 0.030

[a] SNP, the order of all loci is according to physical position on the 3.1-Mb contig

Table 3 Analysis of molecular variance (AMOVA) using 42 markers from ctg954: effect of geographical group

Source of variation	*df*	Variance component	Variation accounted for (%)
Between genepools	1	0.17***	6.06
Between populations within genepools	40	0.57***	20.89
Within populations	334	2.01***	73.05
Total	375	2.75	

*** Significant at $P = 0.001$

6% ($P < 0.001$) reflect differences between European and Asian genepools.

In order to have a complete understanding of the genetic variation at both a contig on chromosome 3B and the whole-genome levels, the 32 SSRs from *ctg954 contig* were used in a comparison with the 70 genome-wide SSRs (Table 4). The mean number of alleles for the 32 SSR loci among the European set of populations was 57.3, ranging from 44 to 68, whereas the corresponding mean value for the Asian set was 60.8 ranging from 48 to 71. The genetic diversity measured by PIC value indicated similar results, with a lower level of PIC value in Europe (mean = 0.242, ranging from 0.122 to 0.323) compared to Asia (mean = 0.265, ranging from 0.164 to 0.339). When considering data from the 70 genome-wide SSRs, the tendency is exactly the same: higher genetic diversity values for both allelic number and PIC value were observed in the Asian populations.

Using the same two sets of SSR data (32 and 70), AMOVA (Table 5) also revealed similar structures of genetic diversity among the 41 populations when comparing the *ctg954* contig from 3B with whole genome: all the different sources of variation are highly significant and, in both case, the most part of variance may be explained by differences within populations (71–73%) while only a small part of the overall variance appears to be the result of difference between Asia and Europe genepools (about 7%).

Population structure

The analysis of population structure of 376 European and Asian accessions based on 70 genome-wide SSRs inferred with the STRUCTURE software led to two groups ($K = 2$) on the basis of Evano criterion. Figure 1 provides a visualization of the structure result. In a second step, due to the specific sample of accession used for this study, all genotypes were assigned to their known original genepools, European and Asian. This assignation appears consistent with the determination of the true value of $K = 2$ for population structure at the molecular level.

When considering the 21 and 20 populations for Europe and Asia, respectively (Table 4), the neighbor-joining tree (Fig. 2) based on a Manhattan distance matrix using information from the 70 SSR markers describes the genetic relationships among the 41 geographical origins. In general, the populations from Europe and Asia clustered together and formed two initial genepools. Interestingly, the Italian population is grouped with materials mainly from South China (Sichuan, Yunnan, Guizhou, Anhui, Hubei, Zhejiang, Fujian, Guangdong). Within the Asian cluster, Chinese accessions were basically divided into North-West China and South-East China, whereas the Nepali, Indian, and Pakistan-Kashmir populations clustered with the Chinese Northwestern provinces of Xinjiang and Tibet-Qinghai. Japanese accessions were genetically close to those from Eastern Chinese provinces, such as Jiangsu and Henan, while Korean varieties appeared to be more distinct, but in the same cluster. Among the European populations, clustering was separated into southeastern Europe (e.g. Romania, Bulgaria, Ukraine-Russia, Yugoslavia), northwestern Europe (e.g. France, Belgium, Netherlands, Sweden), and most Mediterranean countries (Portugal, Greece, Albania and Macedonia), with Spain being somewhat apart and intermediate between Yugoslavia and Germany. In conclusion, both population structure and dendrogram indicate that the European and Asian accessions constitute two genetically independent genepools probably derived from different ancestors.

LD decay across a 3.1-Mb region of chromosome 3B

The most common way to assess decay of LD is to calculate the r^2 value between all pairs of polymorphic marker and to plot these values against the distance between these pairs of markers (Gaut and Long 2003). If we consider the 792 and 799 possible pairs of markers, within the European and the Asian genepools, 251 (29.2%) and 317 (36.8%), showed significant associations at a $P = 0.001$ threshold, respectively. The LD matrix also revealed that 11.7% of the marker pairs in the European pool and 10.9% in the Asian pool showed significant LD values higher than 0.2. Additionally, considering all pairs of r^2 values at $P < 0.001$ thresholds, the European pool had a significantly stronger LD (mean $r^2 = 0.23$) than the Asian genepool (mean $r^2 = 0.18$). Stronger LD can be represented as well by plotting the r^2 value against the physical distance. The average decay of LD throughout the 3.1-Mb region in the European genepool declined to an $r^2 = 0.2$ for genetic intervals as long as 500 kb, but extended up to 2,500 kb with $r^2 = 0.1$ (Fig. 3a). In contrast, the decay of LD in the Asian material showed a different decay pattern of LD, which reached 300 kb with $r^2 = 0.2$, but never decayed to $r^2 = 0.1$ (Fig. 3b). The results suggest that the European genepool had significantly higher mean r^2 value and LD extension ($r^2 = 0.2$) at $P < 0.001$ than that of the Asian.

Comparative analysis of LD with diversity and gene density on ctg954

As LD may be influenced by both genetic diversity and local gene density, we studied the relationship between these parameters in the European and Asian genepools. Local variations of r^2, PIC value and number of genes were estimated within a sliding window along the sequenced contig (with $n = 5$ markers/window). The results (Fig. 4)

Table 4 Name, numbers of accessions, numbers of alleles and PIC values of 21 European and 20 Asian sub-populations used in the study

Gene pool	Geographical area	Population number	Population name	No. of accessions per population	No. of alleles among 32 SSR (ctg954 contig)	PIC value among 32 SSR (ctg954 contig)	No. of alleles among 70 SSR (whole genome)	PIC value among 70 SSR (whole genome)
Europe	Austria	1	AUT	9	63	0.294	190	0.387
	Belgium	2	BEL	9	51	0.156	180	0.359
	Bulgaria	3	BGR	9	60	0.273	232	0.506
	Switzerland	4	CHE	9	51	0.200	209	0.452
	Czech republic	5	CZE	9	61	0.262	186	0.407
	Germany	6	DEU	9	64	0.311	217	0.453
	Spain	7	ESP	9	65	0.279	224	0.459
	Finland	8	FIN	9	62	0.291	208	0.438
	France	9	FRA	9	47	0.151	151	0.313
	England–Ireland	10	GBR–IRL	9	59	0.273	161	0.315
	Greece–Albania–Macedonia	11	GRC–ALB–MAD	9	56	0.240	220	0.473
	Hungary	12	HUN	9	60	0.258	200	0.445
	Italy	13	ITA	9	58	0.276	176	0.374
	The Netherlands	14	NLD	9	54	0.213	137	0.239
	Norway–Denmark	15	NOR–DNK	9	58	0.257	184	0.399
	Poland	16	POL	9	52	0.186	169	0.351
	Portugal	17	PRT	9	63	0.323	241	0.504
	Romania	18	ROM	9	44	0.122	163	0.318
	Sweden	19	SWE	9	68	0.300	179	0.375
	Ukraine–Russia	20	UKR–RUS	9	57	0.230	166	0.316
	Yugoslavia–Croatia	21	YUG–HRV	9	50	0.189	206	0.459
	Mean ± SE				57.3 ± 1.4	0.242 ± 0.012	190.4 ± 6.1	0.397 ± 0.016
	Range				44–68	0.122–0.323	137–241	0.239–0.506
	SD				6.3	0.057	27.9	0.072
Asia	Anhui–Hubei–Zhejiang China	22	AH–HB–ZJ	9	54	0.223	159	0.332
	Beijing China	23	BJ	9	53	0.170	169	0.378
	Fujian–Guangdong, China	24	FJ–GD	8	62	0.305	205	0.468
	Gansu, China	25	GS	9	48	0.164	123	0.229
	Hebei, China	26	HE	8	64	0.312	228	0.487
	Heilongjang, China	27	HL	15	62	0.247	217	0.450
	Henan, China	28	HA	9	58	0.255	231	0.490

Table 4 continued

Gene pool	Geographical area	Population number	Population name	No. of accessions per population	No. of alleles among 32 SSR (ctg954 contig)	PIC value among 32 SSR (ctg954 contig)	No. of alleles among 70 SSR (whole genome)	PIC value among 70 SSR (whole genome)
	Jiangsu, China	29	JS	9	64	0.315	207	0.446
	Jilin–Liaoning, China	30	JL–LN	9	64	0.253	205	0.451
	Shaanxi, China	31	SN	9	59	0.251	215	0.465
	Shandong, China	32	SD	9	54	0.187	192	0.413
	Shanxi, China	33	SX	10	66	0.301	215	0.422
	Sichuan–Yunnan–Guizhou China	34	SC–YN–GZ	9	63	0.315	205	0.451
	Tibet–Qinghai, China	35	XZ–QH	9	64	0.332	243	0.525
	Xinjiang, China	36	XJ	9	69	0.313	246	0.528
	India	37	IND	10	67	0.292	224	0.473
	Japan	38	JPN	9	57	0.255	160	0.321
	Korea	39	KOR	8	60	0.244	243	0.515
	Nepal	40	NPL	9	57	0.226	205	0.378
	Pakistan–Kashmir	41	PAK–KSM	11	71	0.339	261	0.518
	Mean ± SE				60.8 ± 1.3	0.265 ± 0.012	207.7 ± 7.6	0.437 ± 0.017
	Range				48–71	0.164–0.339	123–261	0.229–0.528
	SD				5.8	0.053	33.9	0.077

Table 5 Comparative analysis of molecular variance (AMOVA) using 32 SSR from ctg954 and 70 SSRs from whole genome: effect of geographical group

Source of variation	*df*	AMOVA with 32 SSRs[a]		AMOVA with 70 SSRs[b]	
		Variance component	Variation accounted for (%)	Variance component	Variation accounted for (%)
Between genepools	1	0.11***	6.87	1.19***	6.98
Between populations within genepools	40	0.32***	20.58	3.73***	21.89
Within populations	334	1.13***	72.55	12.12***	71.13
Total	375	1.56		17.04	

*** Significant at $P = 0.001$

[a] SSRs specific for chromosome 3B contig 954

[b] SSRs from whole genome

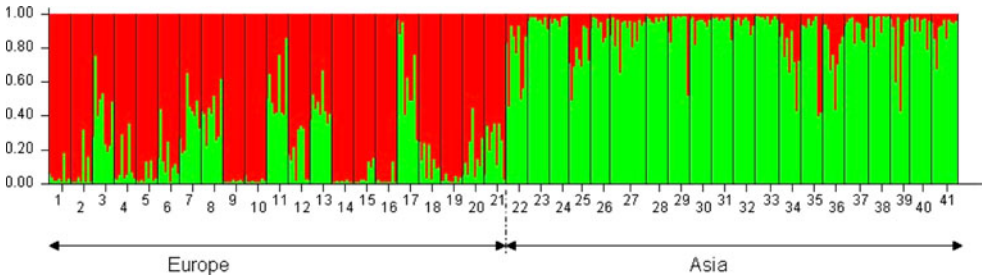

Fig. 1 Population structure analysis of the 376 European and Asian accessions based on 70 genome-wide SSR markers. Each individual is represented by a *thin vertical line*, which is partitioned into $K = 2$ segments that represent the individual estimated membership fraction (%) in K founding groups. Numbers on the x-axis show the population number as indicated in Table 4

showed an uneven distribution of both the PIC and r^2 values in relationship with the density of markers. Overall, very similar trends of variation are observed in diversity and LD between the two sets. However, Asian accessions displayed much higher diversity than European accessions mainly over the right half part of the contig (from position 1.6 Mb to the right end; Fig. 4b). In parallel, LD is lower in Asian accessions from position 2.2 Mb to the right end of the contig (Fig. 4a). And there was an apparent relationship between genetic variation and linkage disequilibrium in the interval from 2.2 to 2.8 Mb, where LD is increasing when genetic diversity is decreasing. The distribution of the number of genes was not homogeneous along the sequence and ranged from 0 to 13 genes per interval (Fig. 4c). When comparing the variation of the gene number with that of LD (Fig. 4a, c), a negative correlation could be observed between these two parameters since intervals with high number of genes generally exhibit lower LD, as for example two regions within interval 1.1–2.5 Mb. Thus, at this level of marker density, these results demonstrate that both genetic diversity and gene content are negatively correlated to LD.

Discussion

Comparison of genetic diversity at different level

At the whole genome scale, our results concerning diversity indices obtained from 70 SSRs (Table 1) exhibit lower level of allelic richness than those reported by Hao et al. (2006) or Roussel et al. (2005) on Asian and European genepools, respectively, Hao et al (2006), analyzing 340 Chinese wheat accessions with 78 genomic SSR loci, detected 13.6 alleles per locus, while Roussel et al. (2005) observed in average 16.4 alleles per locus on 480 European bread wheat accessions with 39 genomic SSR. These differences may be explained by the fact that in our study, a half part of the 70 SSRs used are coming from EST (all the CFE and GPW markers) and are consequently less polymorphic than genomic SSR, as demonstrated by Balfourier et al. (2006) and confirmed by Table 1.

The results at the ctg954 contig level (Table 2) indicated, for CFB SSR, lower values for diversity indices than those obtained at the whole genome scale. Although all the SSR used were genomic SSRs, these differences may be

explained by their motif of repeats. As indicated in supplementary material, most part of them a tri- or tetra-nucleotidic, in contrast with the genomic SSR used at the whole genome scale. As indicated by Balfourier et al. (2006) this could also explain their lower level of diversity.

At the population level, the comparison of the allelic richness and PIC values based on 32 loci located in a 3.1-Mb contig of chromosome 3B (Table 4) demonstrated that Asian populations show higher genetic diversity than the European. The result is similar when considering a genome-wide set of 70 SSR markers.

Results of molecular variance (Tables 3, 5) are of the same magnitude as those presented in previous study for wheat (Roussel et al. 2005): the most part of the diversity is within populations. This demonstrates that the objectives and methods of wheat breeding have a higher impact on recent evolution of genetic diversity than both geographical and environmental conditions. However, even though AMOVA analysis indicated that the differences between Asian and European genepools are weak compared to the within component, these differences are significant and characterized by a higher diversity in Asian material. Finally, the important proportion of private alleles observed at both scale (genome wide and ctg954 contig) confirms that the two genepools differ significantly in term of genetic diversity.

The conclusion of these diversity analyses is that geographic location and environmental conditions similarly influenced genetic variation at the scales of both the specific chromosome region and the whole genome. This is in agreement with Rafalski and Morgante (2004) who indicated that comparative analysis of identical genetic intervals in different populations should eliminate the effect of genomic location, and any differences observed should be due solely to differences between the populations studied.

Geographic structuration in European and Asian wheat germplasm

The neighbor-joining tree (Fig. 2) showed that the 41 populations could be divided into two basic European and Asian genepools, based on analysis with 70 genome-wide SSRs. This is in agreement with previous results (Balfourier et al. 2007; Wang et al. 2007) except for the Italian population that was here clustered with four Chinese populations. This can be explained by the fact that the most part of the material analyzed in the Anhui–Hubei–Zhejiang, Gansu, Fujian–Guangdong and Sichuan–Guizhou–Yunnan populations has an Italian variety as parent, such as cv. Funo, Abbondanza, Mentana or Ardito, as revealed by pedigree records (see http://genbank.vurv.cz/wheat/pedigree/pedigree.asp for wheat pedigree data).

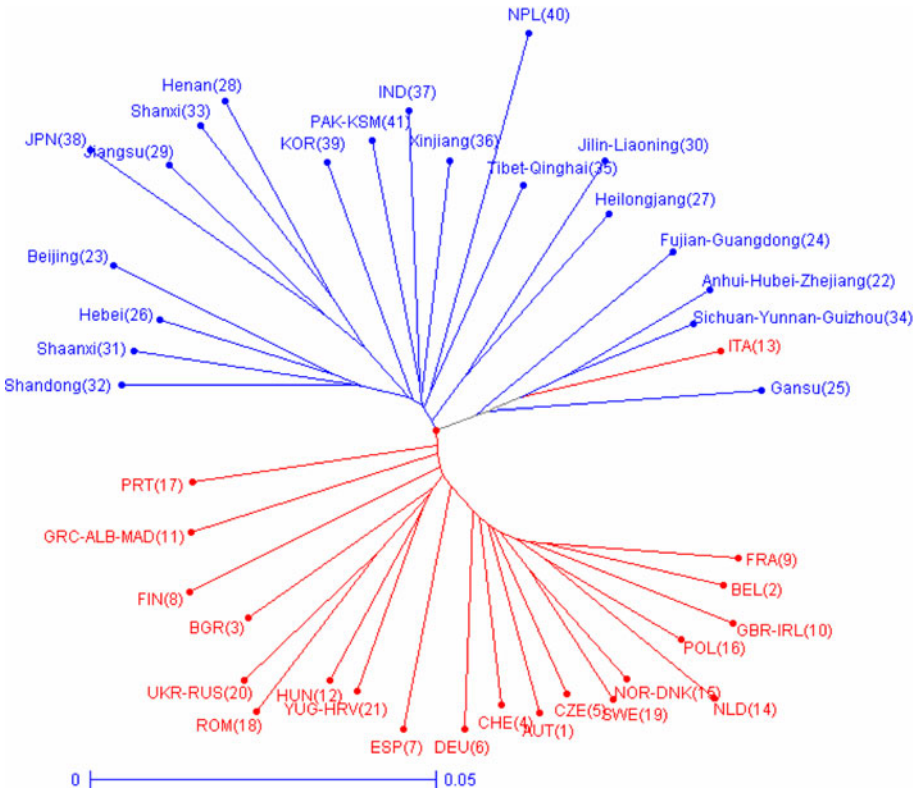

Fig. 2 Neighbor-joining tree of wheat populations based on a Manhattan dissimilarity matrix between 41 populations using 70 genome-wide SSR markers. See Table 4 for abbreviations of name and population number

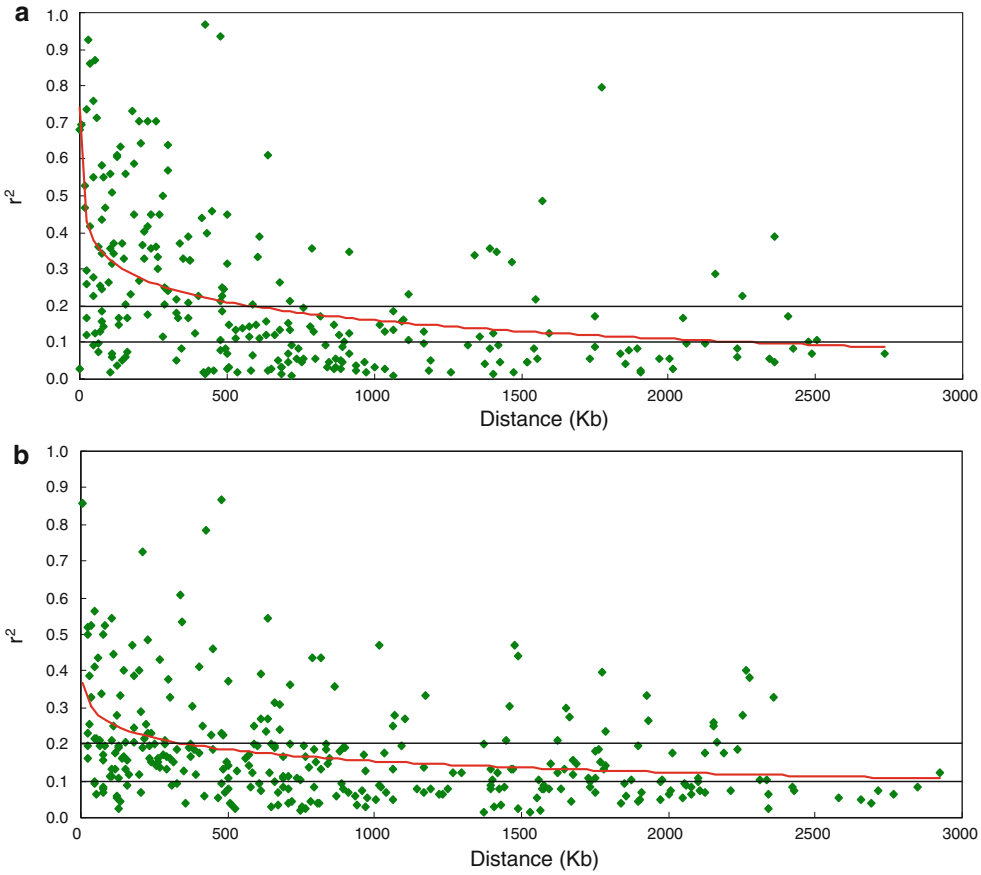

Fig. 3 Plots of significant r^2 value ($P < 0.001$) between pairs of markers in function of the physical distance between these pairs of markers along ctg954 contig, in European (**a**) and Asian (**b**) genepools

To explain the Asian wheat distribution, especially the Chinese genotypes, at least three possible immigration routes were considered by Ghimire et al. (2005): (i) the silk road (from Turkmenistan through Xinjiang to Shaanxi), (ii) the Myanmar route (from Afghanistan through Pakistan, India and Myanmar to Yunnan and Sichuan), and (iii) the Tibetan route (from India through Pamirs or Nepal to Tibet, and then to Shaanxi or Sichuan). We cannot conclude on an imprint of these different routes from the present results, probably because of a limited number of true Chinese landraces in our sample. However, the Xinjiang and Tibet-Qinghai populations formed a large cluster with Nepal, India, and Pakistan–Kashmir, and this clustering would be in favor of the silk or Tibetan road.

The observed distribution of Chinese genotypes (mainly Chinese cultivars) is closely related to the Chinese wheat agro-ecological production zones described by He et al. (2001). For example, the cluster of Beijing, Hebei, Shaanxi and Shandong represents the winter habit zone, whereas facultative wheats typical of the Nepal, India, and Pakistan–Kashmir populations are clustered with the Chinese spring/winter wheat zone of Xinjiang and Tibet–Qinghai. The cluster comprising accessions from the Southeastern provinces (Anhui–Hubei–Zhejiang, Gansu, Fujian–Guangdong and Sichuan–Guizhou–Yunnan) fits well with the autumn-planted spring wheat production zone.

A separation of north-western, south-eastern and Mediterranean groups among the European wheat populations (Fig. 2) is in agreement with earlier reports (Roussel et al. 2005; Balfourier et al. 2007). Thus, our results confirm that the dendrogram was consistent with the distribution of geographic locations and the adaptation of wheat germplasm to the different climatic and environmental conditions of those areas.

Linkage disequilibrium and relationship with genetic diversity and gene density

Both European and Asian wheat genepools have undergone bottlenecks associated with environmental conditions and strong directional selection during their recent histories, which resulted into a non-uniform distribution of genetic diversity and linkage disequilibrium patterns.

The relationship between LD and physical distance has proven to be highly variable in a large range of species, between different types of materials, and also between genes within the same species (Flint-Garcia et al. 2003; Rafalski and Morgante 2004; Yu and Buckler 2006).

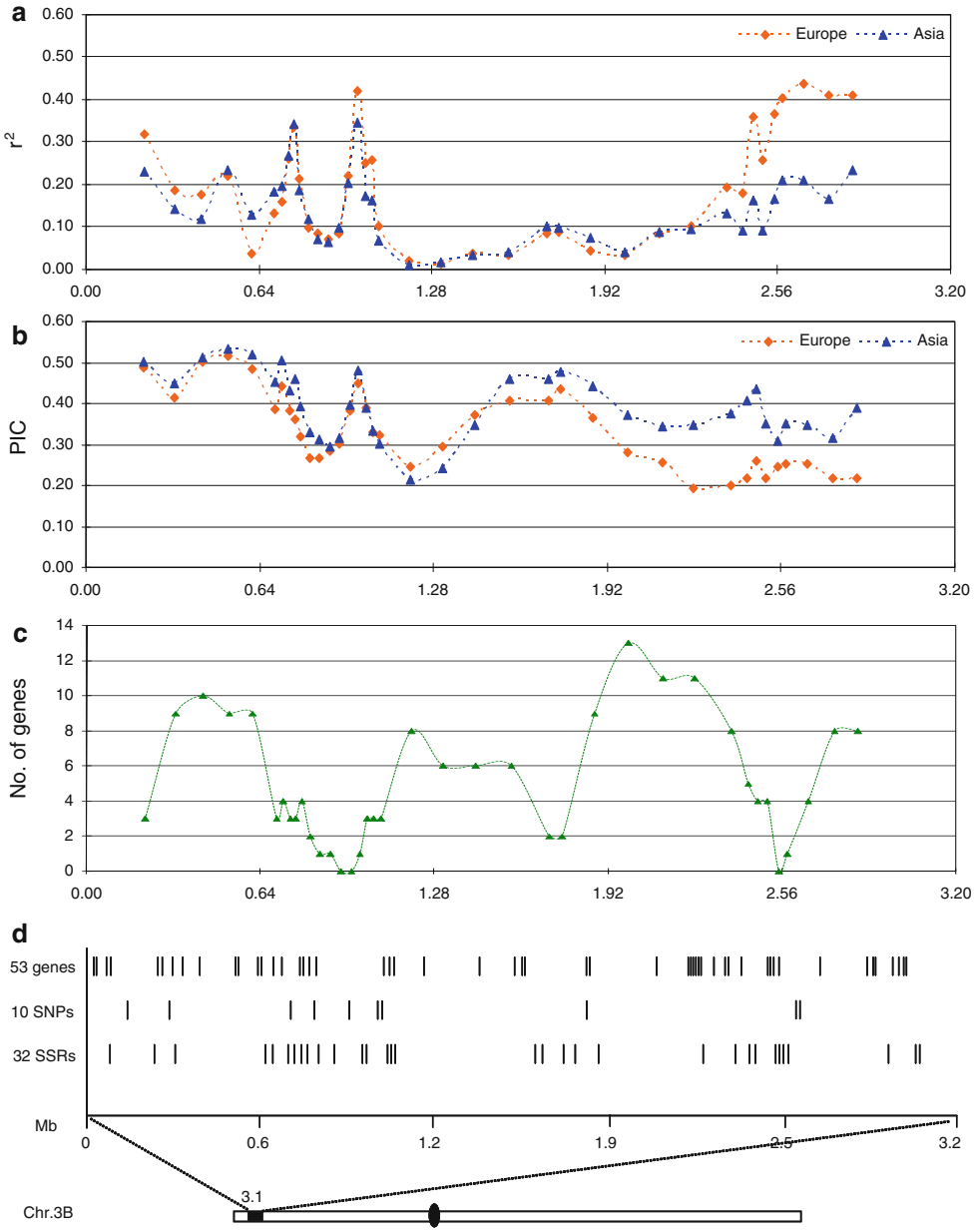

Fig. 4 Distribution of linkage disequilibrium (**a**), genetic diversity (**b**) and number of gene (**c**) along ctg954 contig on chromosome 3B for both the European and Asian genepools using a sliding window approach ($n = 5$ markers per window or interval). Distribution of 32 SSR loci, 10 SNP loci and 53 predicted genes (**d**) along the 3.1 Mb

Recently, for bread wheat species, Horvath et al. (2009) analyzed a highly polymorphic core collection and a more narrow-based breeding material, in order to compare genetic diversity indices and linkage disequilibrium (LD) patterns along the chromosome 3B. These authors demonstrated that LD was weak in both studied materials. Within the core collection, the CIMMYT-ICARDA genepool presented the longest range of LD on the whole chromosome, spreading out to 30 cM contrary to Asian genepool whose LD decreased at 20 cM. Our study also indicate that LD in bread wheat is highly population-dependent, in agreement with the pioneering study of Reich *et al.* (2001) in humans. Within a 3.1-Mb chromosomal region of 3BS, LD extended as long as 500 kb with $r^2 = 0.2$ in the European pool, while it was only 300 kb in the Asian pool, in agreement with Horvath et al. (2009). At the level of four consecutive genes (*gad1, hga2, hga1* and *hga3*) spread along 725 kb on the same 3BS chromosome arm of bread wheat, Ravel et al.(2009) described a low mean r^2 value of LD between *gad1* and *hga3*, confirming the decay of LD within 600–700 kb in this region. However they reported a high level of heterogeneity in the extent of LD, with no LD between *hga2* and *hga3*, separated by 218 kb, and a higher level of LD between *gad1* and *hga2* despite a larger distance (500 kb), suggesting difference of recombination.

In our study, one reasonable explanation for the variation in LD pattern between the two genepools is the

different population histories for the same chromosomal region. The different levels of genetic diversity observed among the populations analyzed suggest varying intensities of selection and breeding systems, producing differences in the amount and extent of LD between the European and Asian germplasm. The influence of domestications on LD and diversity are determined greatly by selection (Nordborg and Tavare 2002). Strong selection at a locus is expected to reduce diversity and increase LD in the surrounding region (Rafalski and Morgante 2004), as demonstrated at the maize *y1* locus by Palaisa et al. (2003). By comparing LD and PIC value levels, our results suggest that intensity of the selection pressure is variable along the 3.1 Mb: for instance, in the interval from 2.2 to 2.8 Mb, which could include the putative *Fhb1* locus related to response to FHB (B. Gill; pers. comm.), LD is increasing when genetic diversity is decreasing with different intensity in both genepools (Fig. 4a, b). In contrast, there was no similar effects in the region between 1.2 to 1.6 Mb, orthologous to the barley *Rph7* locus (Brunner et al. 2003), either due to the absence, in wheat, of selective pressure at this locus or to a lack of markers. Since this contig is included into a multiple disease resistance region, selection pressure caused by presence or absence of pathogens influenced gene diversity at different strengths in the European and Asian genepools.

Concerning the relationship between LD and gene density, as depicted in Rafalski and Morgante (2004), gene-rich regions tend to be subject to low LD probably because of a higher recombination rate close to the genes. In the present study, 41 genes and 12 pseudogenes were predicted from sequencing and annotation analysis. They were unevenly distributed along the sequence and there was a clear relationship between a high number of genes and a low LD at least in two regions, which is consistent with the description by Rafalski and Morgante (2004).

Conclusion and perspectives

The present study is the first detailed description of the level and pattern of genetic diversity and linkage disequilibrium over such a long physical distance in the bread wheat genome i.e. a BAC contig of 3.1 Mb among two contrasting geographical genepools. Our results clearly show that patterns of polymorphism throughout the region are highly nonrandom, and that the extensions of LD are population-dependent. Obviously, the wheat germplasm resources surveyed will be useful for studying LD-based association of complex traits, such as resistance to FHB, a devastating disease which causes significant losses in yield and quality and the accumulation of hazardous mycotoxins in the grain (Zwart et al. 2008). Comparisons of genetic diversity and LD along this *ctg954* contig, which is part of a multiple fungal disease resistance region, indicate effects of apparent divergent artificial selection on putative resistance gene locus in European and Asian wheat genetic resources (Fig. 4a, b). Based on the description of Rafalski (2002), our results indicate that, in the case of putative resistance genes located on this contig, the Asian wheat gene pool used here is more suitable for association mapping than the European one according to apparent linkage disequilibrium patterns. For a candidate gene approach, we will have to develop a higher density of markers. In forthcoming work, we will investigate the phenotypic diversity for fungal disease resistance, such as FHB, among the same wheat lines from Europe and Asia, aiming to discover potential resistance sources for use in wheat breeding and genetic improvement programs.

Acknowledgments This work was supported by a Project of Advanced Research (PRA 005) between the French National Institute for Agronomical Research (INRA) and the Chinese Academy of Agriculture Science. It was also funded by the EXEGESE-BLE and Genoplante-SMART projects from the French National Research Agency. The authors are grateful to the MRI from INRA for providing a grant and funding for the experiment. They also gratefully appreciate Prof. Robert A McIntosh of the University of Sydney for his help on English editing.

References

Able JA, Langridge P, Milligan AS (2007) Capturing diversity in the cereals: many options but little promiscuity. Trends Plant Sci 12:71–79

Balfourier F, Ravel C, Bochard AM, Exbrayat Vinson F, Boutet G, Sourdille P, Dufour F, Charmet G (2006) Développement utilisation et comparaison de différents types de marqueurs pour étudier la diversité parmi une collection de blé tendre. Les Actes du BRG 6:129–144

Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, Koenig J, Ravel C, Mitrofanova O, Beckert M, Charmet G (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. Theor Appl Genet 114:1265–1275

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Brunner S, Keller B, Feuillet C (2003) A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the *Pph7* locus. Genetics 164:673–683

Chao S, Zhang WJ, Dubcovsky J, Sorrells M (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. wheat (*Triticum aestivum* L.) germplasm representing different market classes. Crop Sci 47:1018–1030

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier MC, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell (in press)

Dreher K, Khairallah M, Ribaut J-M, Morris M (2003) Money matters (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. Mol Breed 11:221–234

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Feldman M (2000) Origin of cultivated wheat. In: Bonjean AP, Angus JW (eds) The world wheat book, a history of wheat breeding. TEC and DOC, London, pp 3–56

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Frankham R (2005) Genetics and extinction. Biol Conserv 126:131–140

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1502–1506

Ghimire SK, Akashi Y, Maitani C, Nakanishi M, Kato K (2005) Genetic diversity and geographical differentiation in Asian common wheat (*Triticum aestivum* L.), revealed by the analysis of peroxidase and esterase isozymes. Breed Sci 55:175–185

Hao CY, Wang LF, Zhang XY, You GX, Dong YS, Jia JZ, Liu X, Shang XW, Liu SC, Cao YS (2006) Genetic diversity in Chinese modern wheat varieties revealed by microsatellite markers. Sci China C Life Sci 49:218–226

Hao CY, Dong YS, Wang LF, You GX, Zhang HN, Ge HM, Jia JZ, Zhang XY (2008) Genetic diversity and construction of core collection in Chinese wheat genetic resources. Chin Sci Bull 53:1518–1526

He ZH, Rajaram S, Xin ZY, Huang GZ (2001) A history of wheat breeding in China. CIMMYT, Mexico, D.F.

Horvath A, Didier A, Koenig J, Exbrayat F, Charmet G, Balfourier F (2009) Analysis of diversity and linkage disequilibrium along chromosome 3B of bread wheat (*Triticum aestivum* L.). Theor Appl Genet. doi:10.1007/s00122-1158-8

Huang XQ, Borner A, Röder MS, Ganal MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 105:699–707

Kilian B, Özkan H, Pozzi C, Salamini F (2009) Domestication of the Triticeae in the Fertile Crescent. In: Feuillet C, Muehlbauer G (eds) Genetics and genomics of the Triticeae plant genetics and genomics: crops and models 7. Springer, Berlin, pp 81–119

Liu K, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–289

Mitrofanova OP, Strelchenko PP, Konarev AV, Balfourier F (2009) Genetic differentiation of hexaploid wheat inferred from analysis of microsatellite loci. Russ J Genet 45(11):1351–1359

Morris M, Dreher K, Ribaut JM, Khairallah M (2003) Money matters (I): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. Mol Breed 11:235–247

Nei M (1973) Analysis of gene diversity in subdivided populations. PNAS 70:3321–3323

Nesbitt M (2001) Wheat evolution: integrating archaeological and biological evidence. The Linnean (special issue) 3:37–59

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy F, Bernard M, Sourdille P (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109:800–805

Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18:83–90

Palaisa K, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15:1795–1806

Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, Feuillet C (2006) Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. Plant J 48:463–474

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P et al (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101–104

Paux E, Faure S, Choulet F, Roger D, Gauthier V, Martinant JP, Sourdille P, Balfourier F, Le Paslier MC, Chauveau A, Cakir M, Gandon B, Feuillet C (2010) Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnol J 8:196–210

Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Enfield, Science Publishers, Montpellier, pp 43–76

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure from multilocus genotype data. Genetics 155:945–959

Rafalski A (2002) Application of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet 20:103–111

Raquin AL, Brabant P, Rhone B, Balfourier F, Leroy P, Goldringer I (2008) Soft selective sweep near a gene that increases plant height in wheat. Mol Ecol 17:741–756

Ravel C, Choulet F, Dardevet M, Exbrayat-Vinson F, Sourdille P, Balfourier F (2009) Linkage disequilibrium at different scales on 3B chromosome of bread wheat. In: Proceedings of ITMI international conference, 31 August–4 September, Clermont-Ferrand, France, p 132

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411:199–204

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganal MW (2002) Construction and analysis of a microsatellite-based database of European wheat varieties. Theor Appl Genet 106:67–73

Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F (2005) SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor Appl Genet 111:162–170

Saintenac C, Falque M, Martin O, Paux E, Feuillet C, Sourdille P (2009) Detailed recombination studies along chromosome 3B provide new insight into crossover distribution in wheat. Genetics 181:393–403

Schlumbaum A, Neuhaus JM, Jacomet S (1998) Coexistence of tetraploid and hexaploid naked wheat in a neolithic lake dwelling of central Europe. Evidence from morphology and ancient DNA. J Archaeol Sci 25:1111–1118

Somers DJ, Banks T, DePauw R, Fox S, Clarke J, Pozniak C, McCartney C (2007) Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome 50:557–567

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RELF mapping in plant breeding: new tools for an old science. Biotechniques 7:257–264

Väli Ü, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations. Mol Ecol 17:3808–3817

Wang LF, Balfourier F, Hao CY, Exbrayat-Vinson F, Dong YC, Ge HM, Zhang XY (2007) Comparison of genetic diversity level between European and East-Asian wheat collections using SSR markers. Sci Agric Sin 40:2667–2678

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733

Yu JM, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:155–160

Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. Theor Appl Genet 111:677–687

Zhang ZW, Bradbury PJ, Kroon DE, Casstevens TM, Buckler ES (2006) TASSEL 2.0: a software package for association and diversity analyses in plants and animals (http://www.maizegenetics.net). In: Plant and animal genomes XIV conference, Poster P956/CP012, San Diego, USA

Zwart RS, Muylle H, Bockstaele EV, Roldán-Ruiz I (2008) Evaluation of genetic diversity of Fusarium head blight resistance in European winter wheat. Theor Appl Genet 117:813–828